



JISTech (Journal of Islamic Science and Technology)

JISTech, 6(1), 40-51, Januari-Juni 2021

ISSN: 2528-5718

<http://jurnal.uinsu.ac.id/index.php/jistech>

IMPLEMENTASI ALGORITMA C 5.0 DALAM PREDIKSI STOK BARANG BERDASARKAN PENJUALAN BAHAN PERTANIAN CV. MITRA SEJATI

Siti Mawaddah

Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: sitimawaddahtanjung@gmail.com¹, ali_ikhwan@uinsu.ac.id²,
muhammadeddiirawan@uinsu.ac.id³

ABSTRACT

CV. Mitra Karya Sejati was founded in 1998 by Tardi Sutanto as a trading company in the field of agricultural materials, such as multipurpose fertilizers, pesticides and seeds. CV. Mitra Karya Sejati is often in short supply due to the increasing demand for consumer goods and even some items are not sold, so it is difficult to predict the stock of goods in the warehouse so far based on sales data which is used only as data to report monthly sales results which will then not be used again. . Companies need a sales forecasting system to help you plan stock items in the future. The application of C 5.0 to data mining can make it easier for companies to estimate stock of goods based on sales of agricultural materials. The Data Mining application is built using the PHP programming language and MySQL database to reduce inventory planning losses. Through the proposed system can help companies estimate the stock of goods to be provided to minimize losses.

Keywords: *C 5.0, Data Mining, Prediction.*

PENDAHULUAN

Pesatnya perkembangan teknologi pada saat ini memudahkan masyarakat untuk mendapatkan berbagai macam informasi. Hampir semua masyarakat Indonesia baik untuk anak-anak, remaja atau dewasa setiap hari terlibat dengan teknologi. Peran teknologi untuk saat ini dianggap mampu menggantikan kegiatan manusia. Dengan berkembang pesatnya

teknologi saat ini maka penyajian informasi secara efektif dan efisien semakin dibutuhkan. Di mana penyajian informasi seharusnya sudah beralih dari sistem lama pencatatan manual menjadi sistem komputerisasi [1].

CV. Mitra Karya Sejati Jl. Menggala Sakti Km.24, Kec. Tanah Putih, Kab. Rokan Hilir, Riau adalah perusahaan dagang yang menawarkan alat dan bahan pertanian dengan harga murah. CV. Mitra Karya Sejati menggunakan sistem informasi penjualan untuk mendukung kinerja perusahaan dalam layanan informasi penjualan. Ketersediaan produk pertanian dengan harga murah menjadi tantangan tersendiri untuk menjaga kestabilan stok barang di gudang. Walaupun kelihatannya mudah, namun mengelola persediaan bukanlah sesuatu yang bisa dianggap remeh, karena jika persediaan terlalu banyak, risiko kerusakan barang akan meningkat, dan sebaliknya, jika persediaan sedikit, risiko kekurangan barang akan lebih tinggi dari persediaan dan dapat menunda keuntungan serta membuat konsumen kecewa.

Penjualan bahan pertanian telah meningkat dalam beberapa bulan terakhir, sehingga ketersediaan barang di gudang sering kekurangan karena meningkatnya permintaan barang konsumsi dan fakta bahwa beberapa barang bahkan tidak terjual. Seperti yang terjadi pada bulan Juni hingga bulan September pada tahun 2021 di mana mengalami kekurangan stok barang pada pupuk Meroke dan kerusakan pada pupuk CU Daun sehingga mengakibatkan kerugian, hal ini disebabkan karena banyaknya pemesanan pada pupuk tertentu di perusahaan pada bulan Juni hingga September, sehingga ada kesulitan untuk prediksi stok barang pada gudang penyimpanan. Pengumuman ini sangat berguna dalam menentukan berapa banyak produk yang akan dikirimkan bulan depan. Masalah yang sering dihadapi perusahaan adalah tingkat akurasi dalam memprediksi stok komoditas masa depan berdasarkan data penjualan sebelumnya. Prediksi tersebut berpengaruh besar dalam menentukan produk mana yang laku dan mana yang tidak. Penelitian ini didasarkan pada data penjualan CV. Mitra Karya Sejati yang menggunakan data penjualan bulanan hanya sebagai data untuk melaporkan hasil penjualan bulanan, kemudian tidak akan

digunakan lagi.

Data mining dapat digunakan untuk mengekstrak informasi dari *big data* untuk memperoleh informasi yang dapat digunakan untuk memprediksi, mengklasifikasi, dan memperkirakan. Metode Algoritma C 5.0 Data mining adalah salah satu algoritma klasifikasi yang secara khusus dapat diterapkan pada teknisi *Decision Tree*. Algoritma C 5.0 merupakan perbaikan dari algoritma ID3 dan C 4.5 [2].

Berdasarkan fenomena dan permasalahan tersebut, penelitian ini bertujuan untuk memudahkan pihak penyedia stok barang yang berada di CV. Mitra Karya Sejati dalam merencanakan persediaan barang dan memudahkan perusahaan untuk mengetahui produk pupuk yang perlu dipasok dan banyak diminati oleh konsumen menggunakan algoritma C 5.0.

LANDASAN TEORI

1. Algoritma C 5.0

Algoritma C 5.0 adalah salah satu *data mining* yang khususnya diterapkan pada teknik *Decision Tree*. C 5.0 merupakan penyempurnaan dari algoritma sebelumnya yaitu, ID3 dan C 4.5 yang diperkenalkan terlebih dahulu oleh J.Ross Quinlan pada tahun 1987. Dalam algoritma ini pemilihan atribut yang akan diproses menggunakan *Information Gain*. Memilih atribut untuk pemecah objek dalam beberapa kelas harus dipilih dari atribut yang menghasilkan *Information Gain* paling besar. Atribut dengan nilai *Information Gain* tertinggi akan dipilih sebagai *parent* untuk *node* selanjutnya [3].

2. Prediksi

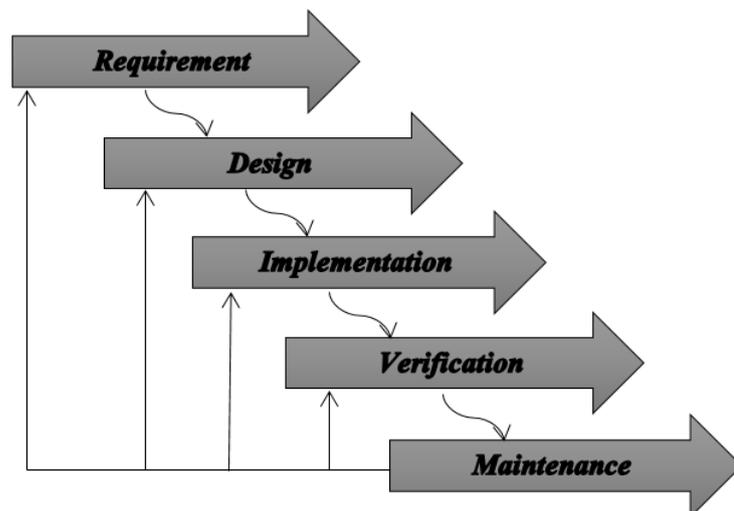
Prediksi (peramalan) adalah usaha menduga atau memperkirakan sesuatu yang akan terjadi di waktu mendatang dengan memanfaatkan berbagai informasi yang relevan pada waktu-waktu sebelumnya (*historis*) melalui suatu metode ilmiah. Tujuan dari prediksi adalah mendapatkan informasi apa yang akan terjadi di masa datang dengan probabilitas kejadian terbesar [4].

3. Stok Barang

Persediaan merupakan salah satu aspek yang perlu diperhatikan oleh setiap perusahaan. Hal ini dipicu adanya pengeluaran yang ditimbulkan dalam aktivitas persediaan, atau sering disebut sebagai biaya persediaan. Idealnya, kontrol persediaan barang dapat dilakukan lebih mudah jika ada sistem pendukung yang membantu proses pengelolaan. Namun pada kenyataannya kegiatan ini masih dilakukan manual [5].

METODE PENGEMBANGAN SISTEM

Penelitian ini menggunakan metode *waterfall* sebagai metode pengembangan sistem yang dibangun. Tahapan metode *waterfall* digambarkan pada gambar 1 di bawah ini.



Gambar 1. Tahapan Metode *Waterfall* [6]

Tahapan dalam *Waterfall* adalah sebagai berikut:

1. Requirement

Pada tahap ini, pengembang sistem perlu berkomunikasi dengan pengguna untuk memahami perangkat lunak dan keterbatasan perangkat lunak yang mereka harapkan. Informasi dapat diperoleh melalui wawancara, diskusi atau *live survey*. Informasi tersebut dianalisis untuk mendapatkan data yang dibutuhkan pengguna.

2. Design

Pada tahap ini, pengembang membuat desain sistem yang dapat membantu menentukan perangkat keras dan kebutuhan sistem dan juga

membantu menentukan arsitektur sistem secara keseluruhan..

3. *Implementation*

Pada fase ini, sistem pertama kali dikembangkan menjadi program-program kecil yang disebut unit, yang diintegrasikan ke dalam fase-fase berikutnya. Setiap unit dikembangkan dan diuji untuk fitur yang disebut unit test.

4. *Verification*

Pada fase ini, sistem diperiksa dan diuji untuk melihat apakah sistem sepenuhnya atau sebagian memenuhi persyaratan sistem, pengujian dapat diklasifikasikan sebagai pengujian unit (dilakukan pada modul kode tertentu), pengujian sistem (untuk melihat bagaimana sistem bereaksi ketika semua modul terintegrasi) dan tes penerimaan (dengan atau atas nama pelanggan untuk melihat apakah semua kebutuhan pelanggan terpenuhi).

5. *Maintenance*

Ini adalah langkah terakhir dalam metode *Waterfall*. Perangkat lunak yang telah selesai dijalankan dan dipelihara. Pemeliharaan memperbaiki *bug* yang tidak ditemukan di langkah sebelumnya [6].

HASIL DAN PEMBAHASAN

1. Penerapan Algoritma C 5.0

Pada penelitian ini terdapat sebuah data set dengan jumlah sebanyak 1356 *record*, yang mana data tersebut memiliki 6 atribut dan 1 kelas.

a. *Preprocessing Data* dan Menghitung *Entropy*

Langkah awal adalah *processing* data, yaitu proses *cleaning* data yang kemudian akan dilakukan perhitungan *Entropy* dan *Gain*. Rumus perhitungan *Entropy* dituliskan pada persamaan (1) [7].

$$Entropy(S) = \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

Keterangan:

S : himpunan

A : atribut

n : jumlah partisi S

pi : proporsi dari Si terhadap S

Dengan rumus pada persamaan (1) tersebut mendapatkan hasil *entropi* seluruh kelas adalah 0,847618343.

b. Menghitung Gain

Tahap kedua adalah melakukan perhitungan *Gain* yang dituliskan pada persamaan (2) berikut ini [8].

$$Gain(S,A) = Entropy(S) \sum_{j=1}^Y \frac{D_j}{D} * Entropy(S_j) \tag{2}$$

c. Menghitung *Split Information*

Selanjutnya dilakukan perhitungan *split information* yang digunakan untuk memilih sejumlah atribut, dengan rumusnya ada pada persamaan (3) dibawah ini [9].

$$SplitInfo(S,A) = \sum_i^c = 1 \frac{S_i}{S} \text{Log}_2 \frac{S_i}{S} \tag{3}$$

Di mana S adalah Jumlah Data Sampel dan Si adalah jumlah masing-masing pada setiap atribut.

d. Menghitung Nilai *Gain Ratio*

Setelah mendapat hasil *split information* tahap selanjutnya adalah mengitung *Gain Ratio* yang dituliskan pada rumus persamaan (4) berikut ini [10].

$$Gain Ratio = \frac{Gain(S,A)}{Split Info(S,A)} \tag{4}$$

Di bawah ini adalah hasil perhitungan masing-masing atribut menggunakan persamaan (1), (2), (3), dan (4).

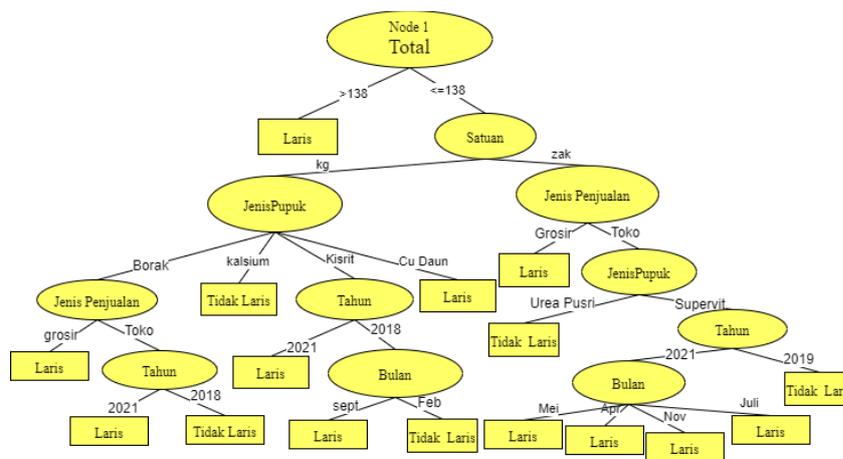
Tabel 1. *Tabel Perhitungan Gain Ratio*

Atribut	<i>Informasi Gain</i>	<i>Split Info</i>	<i>Gain Ratio</i>	Peringkat
Nama Pupuk	0,037737	5,26530907	0,007167186	
Jenis	0,001894849	0,84761834		
Penjualan			0,002235498	
Satuan	0,0000312	0,81244488	0,0000384	
Bulan	0,063005643	3,28935216	0,019154423	
Tahun	0,003255712	1,99656532	0,001630656	
Total	0,847618343	7,88465597	0,107502261	1

Dari tabel di atas, dapat disimpulkan bahwa *Gain Ratio* tertinggi terletak pada atribut total dengan nilai 0,10750226 sehingga dapat diambil menjadi patokan dalam *node* pohon keputusan.

e. Pohon Keputusan

Menentukan pohon keputusan yaitu mencari nilai *Gain Ratio* tertinggi dari semua atribut, kemudian membuat *node* akar pohon keputusan berdasarkan nilai *Gain Ratio* tertinggi. Dari hasil tersebut sudah didapatkan *node* akar atau simpul pertama pohon keputusan. Gambar 2 di bawah ini adalah gambar dari pohon keputusan.



Gambar 2. Pohon Keputusan

Pada gambar 2 bisa dilihat bahwa *gain ratio* tertinggi adalah total, jika penjualan yang totalnya >138 maka penjualan itu adalah laris, dan jika total <= 138 maka keputusannya belum ada karena harus dihitung lagi node atau akarnya.

f. Confusion Matrix

Confusion matrix merupakan sebuah metode untuk evaluasi yang menggunakan tabel *matrix*. Hasil evaluasi dengan menggunakan *confusion matrix* menghasilkan akurasi, serta laju *error*. Akurasi menyatakan jumlah data yang diklasifikasi benar setelah dilakukan proses pengujian, sedangkan laju *error* digunakan untuk menghitung kesalahan identifikasi. Untuk menghitung akurasi adalah:

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+FN} \times 100 \tag{5}$$

Di mana *TP* adalah *True Positive* yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem, *TN* adalah *True Negative* yaitu

jumlah data negatif yang terklasifikasi dengan benar oleh sistem, *FN False Negative* yaitu jumlah negatif namun terklasifikasi salah oleh sistem dan *FP* adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem [11]. Laju *error* dapat dihitung sebagai berikut:

$$\text{Laju error} = \frac{\text{Jumlah data diidentifikasi salah}}{\text{Jumlah seluruh data}} \times 100 \tag{6}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{8}$$

Di mana:

TP (True Positive) = Jumlah data Aktual='Laris' dan Prediksi ='Laris'

FN (False Negative) = Jumlah data Aktual='Laris' dan Prediksi='Tidak Laris'

TN (True Negative) = Jumlah data Aktual='Tidak Laris' dan Prediksi='Laris'

FP (False Positive) = Jumlah data Aktual='Tidak Laris' dan Prediksi='Tidak

Tabel 2. Confusion Matrix

	Diidentifikasi Laris	Diidentifikasi Tidak laris
Data Uji Laris	193	0
Data Uji Tidak Laris	0	79

Dari tabel 2 di atas, dapat diukur tingkat akurasi dari klasifikasi sebagai berikut:

$$TP = 193, FN = 0, TN = 79, FP = 0$$

$$\text{Akurasi} = ((TP + TN) / (TP+FP+TN+FN)) \times 100\%$$

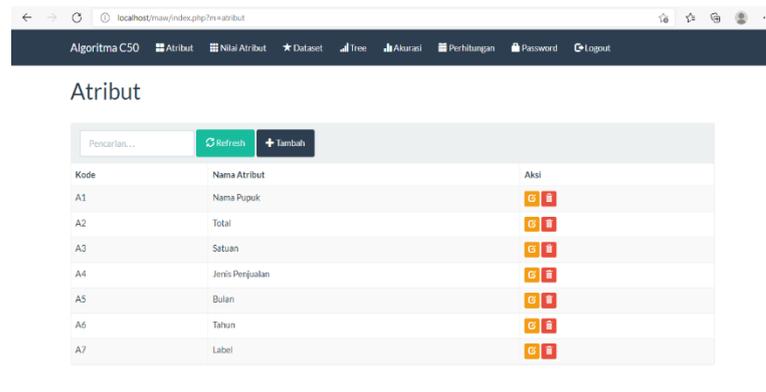
$$\text{Akurasi} = ((193+79) / (193+0+79+0)) \times 100\%$$

$$\text{Akurasi} = 100\%$$

IMPLEMENTASI SISTEM

1. Halaman Atribut

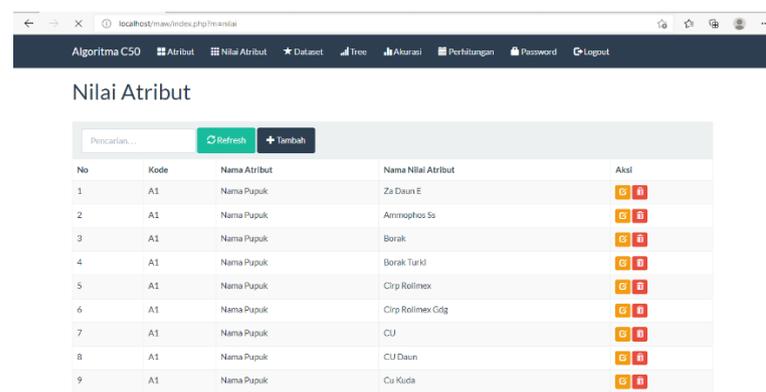
Halaman ini berfungsi untuk mengelola atribut seperti menambah, mengedit, serta menghapus. Gambar 3 di bawah ini adalah tampilan halaman atribut.



Gambar 3. Halaman Atribut

2. Halaman Nilai Atribut

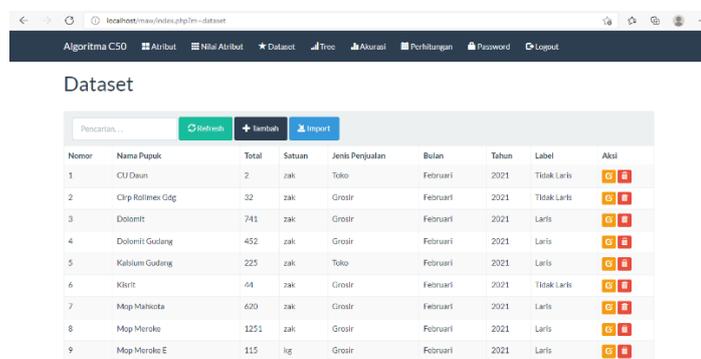
Halaman ini merupakan halaman yang menampilkan nilai atribut dan dapat mengelola, menghapus, dan edit data nilai atribut. Berikut adalah tampilan nilai atribut:



Gambar 4. Halaman Nilai Atribut

3. Halaman *Dataset*

Halaman *dataset* merupakan halaman yang dapat mengelola, mengedit, dan menghapus *dataset*. Di bawah ini adalah tampilan halaman *dataset*:



Gambar 5. Halaman *Dataset*

4. Tampilan Hasil Perhitungan

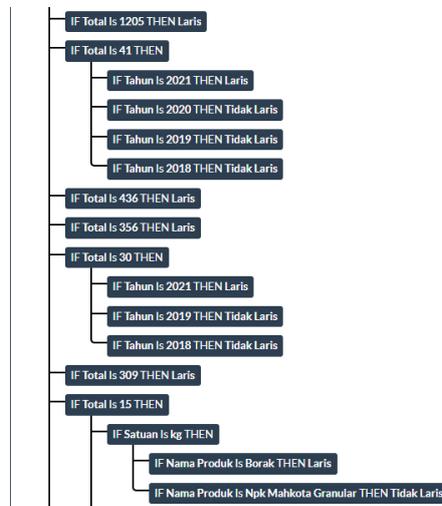
Halaman ini adalah halaman perhitungan metode algoritma *C 5.0* seperti *entropy*, *gain*, *splitInfo*, dan *gain ratio*. Halaman hasil perhitungan dapat dilihat pada gambar 6 di bawah ini.

No.	Jenis Atribut	Nama Atribut	Nilai	Total	Entropy	Gain	Split Info	Gain Ratio
1	Nama Produk	CU Daun	30	1356	1	0.038	5.273	0.007
2	Nama Produk	Cirp Rollmex Gdg	31	1356	0.771			
3	Nama Produk	Dolomit	53	1356	0.657			
4	Nama Produk	Dolomit Gudang	35	1356	0.513			
5	Nama Produk	Kalsium Gudang	42	1356	0.792			
6	Nama Produk	Kisrit	9	1356	0.764			

Gambar 6. Tampilan Hasil Perhitungan

5. Tampilan Pohon Keputusan

Halaman pohon keputusan merupakan salah satu hasil dari metode *C 5.0* yang menjadikan klasifikasi berbentuk pohon keputusan. Berikut tampilan dari halaman keputusan:



Gambar 7. Tampilan Pohon Keputusan

6. Tampilan Hasil Akurasi

Dari data uji didapat nilai akurasi data uji laris diidentifikasi laris *C 5.0* berjumlah 193, diidentifikasi tidak laris *C 5.0* berjumlah 0. Kemudian data uji tidak laris diidentifikasi laris *C 5.0* berjumlah 0, dan diidentifikasi tidak laris *C 5.0* berjumlah 79. Maka memiliki nilai akurasi 100%. Dapat dilihat pada gambar di bawah ini:

Confusion Matrix, Akurasi, Precision dan Recall		
1. Confusion Matrix		
	Diidentifikasi Laris C.50	Diidentifikasi Tidak Laris C.50
Keputusan data uji Laris	193	0
Keputusan data uji Tidak Laris	0	79

2. Akurasi = 100%
3. Precision = 100%
4. Recall = 100%

Gambar 8. Tampilan Hasil Akurasi

KESIMPULAN

Penelitian ini berhasil mengimplementasikan Algoritma C 5.0 pada perancangan aplikasi *data mining* untuk mengetahui perencanaan stok barang berdasarkan penjualan bahan pertanian pada CV. Mitra Karya Sejati. Dari hasil penerapan Algoritma C 5.0 didapatkan node akar atau simpul pertama pohon keputusannya yaitu total dari penjualan. Penerapan Algoritma C 5.0 pada aplikasi *data mining* ini dapat membantu pihak CV. Mitra Karya Sejati dalam memutuskan stok barang, yaitu dengan mengetahui berapa besar penjualan yang laris dan tidak laris. Keseluruhan proses pada aplikasi *data mining* yang dirancang telah berhasil dijalankan dan dapat difungsikan sesuai tujuan yang diinginkan.

REFERENSI

- [1] M. Danuri, "Development and Transformation of Digital Technology," *Infokam*, vol. XV, no. II, pp. 116–123, 2019.
- [2] A. W. Wijayanti, "Analisis Hasil Implementasi Data Mining Menggunakan Algoritma Apriori pada Apotek," *J. Edukasi dan Penelit. Inform.*, vol. 3, no. 1, pp. 60–64, 2017, doi: 10.26418/jp.v3i1.19534.
- [3] G. A. Darmawan, Y. Susanti, and Siswanto, "Implementasi Data Mining Menggunakan Algoritme C5.0 Pada Data Kelulusan Mahasiswa S1 Universitas Sebelas Maret Surakarta," *Univ. Sebel. Maret*, vol. 1, no. 1, pp. 1–8, 2018.
- [4] L. W. Ningsih and A. A. Nababan, "Perancangan Aplikasi Untuk

- Prediksi Pengeluaran Dana Desa Dengan Metode Algoritma Moving Average,” *JIKOMSI (Jurnal Ilmu Komput. dan Sist. Informas)*, vol. 3, no. 3, pp. 103–108, 2021.
- [5] L. P. A. Prapitasari and N. K. D. A. Jayanti, “Penerapan Metode EOQ (Economic Order Quantity) Pada Peramalan Stok Barang,” *Konf. Nas. Sist. Inform.*, pp. 9–10, 2017.
- [6] A. A. Wahid, “Analisis Metode Waterfall Untuk Pengembangan Sistem Informasi,” *J. Ilmu-ilmu Inform. dan Manaj. STMIK*, pp. 1–5, 2020.
- [7] M. S. Sungkar and M. T. Qurohman, “Penerapan Algoritma C5.0 Untuk Prediksi Kelulusan Pembelajaran Mahasiswa Pada Matakuliah Arsitektur Sistem Komputer,” *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1166, 2021, doi: 10.30865/mib.v5i3.3116.
- [8] N. H. Harani and F. S. Damayanti, “Implementasi Algoritma C5.0 Untuk Menentukan Pelanggan Potensial Di Kantor Pos Cimahi,” *J. SITECH Sist. Inf. dan Teknol.*, vol. 4, no. 1, pp. 69–76, 2021, doi: 10.24176/sitech.v4i1.6281.
- [9] M. R. Santoso and P. Musa, “Rekomendasi Kesehatan Janin Dengan Penerapan Algoritma C5.0 Menggunakan Classifying Cardiotocography Dataset,” *J. Simantec*, vol. 9, no. 2, pp. 65–76, 2021, doi: 10.21107/simantec.v9i2.10730.
- [10] F. N. Umma, B. Warsito, and D. A. I. Maruddani, “Klasifikasi Status Kemiskinan Rumah Tangga Dengan Algoritma C5.0 Di Kabupaten Pematang,” *J. Gaussian*, vol. 10, no. 2, pp. 221–229, 2021, doi: 10.14710/j.gauss.v10i2.29934.
- [11] R. N. Amalda, N. Millah, and I. Fitria, “Implementasi Algoritma C5.0 dalam Menganalisa Kelayakan penerima Keringanan UKT Mahasiswa ITK,” *Teorema Teor. dan Ris. Mat.*, vol. 7, no. 1, pp. 101–116, 2022.